

## 言語能力測定システムにおける負荷軽減手法の提案

## Proposal of Burden Reduction Method for Measurement System of Language Ability

長田 颯斗<sup>†</sup> 荒牧 英治<sup>††</sup> 宮部 真衣<sup>†††</sup>  
Hayato Choda Eiji Aramaki Mai Miyabe

## 1. はじめに

近年、客観的に言語能力を測る重要性が様々な場面で高まっている。例えば、北村は、「言語能力を測ることは、早期発見や療養に役立つ可能性がある」と述べている [1]。そのため、言語能力測定は認知症の早期発見において有用であると考えられる。また、現在 26 万人を超える外国人留学生が日本語を学んでおり [2]、外国人留学生を採用する企業も近年増加している [3] が、採用選考の場において外国人留学生の日本語能力の評価に課題を抱えている企業もある<sup>1</sup>。これは、会話の場において、面接官による判断の揺れがしばしば客観的な評価を困難にするためであり、客観的な言語能力測定が求められている。

従来、言語能力の測定には、訓練された専門家が必要となるため、手軽に測定することが難しい。我々はこれまでに、言語能力測定システム「言秤」を開発した [4]。言秤では、音声発話により入力された内容をもとに、6 つの言語能力指標により利用者の言語能力を測ることができる。しかし、言秤ではパソコンに向かって直接発話する必要があることから、抵抗感を与える可能性があり、利用に対する敷居の高さが課題になっている。

そこで、本研究では、言語能力測定の負担軽減を目的とした言語能力測定システム「言秤 WEB」を開発した。本稿では、提案手法を実装し、システムによる測定負担に対する効果について述べる。

## 2. 関連研究

言語能力測定は、高齢者や外国人留学生だけでなく、子供を対象とすることも求められている。Longobardi らは、言語が発達している子供は言語発達が遅い子と比較して社会的能力が低く [5]、言語能力は社会的能力を予測することができることを明らかにした。また、Yoshinaga-Itano らは、早期発見群 (6 ヶ月以前) の難聴者と後期発見群 (6 ヶ月以後) の難聴者で比較した結果、早期発見群の難聴者の方が言語能力が高く [6]、早期に難聴を確認することは言語能力の向上に関係があることを示した。本システムは、利用者自身が測定するだけでなく、第三者を介して言語能力測定を行うことができる。例えば、パソコン操作などができない子供でも、保護者がシステムを利用することで子供の言語能力を測定できる。

また、言語能力はコミュニケーション能力と深い関係がある。鳥飼は、コミュニケーション能力は、コミュニケー

ションを実際に行う上で必要になる能力の基板として、「言葉の知識」だけでなく、「知識体系全般」が主要な要素であると述べている [7]。本システムでは、文章の丁寧さや構成能力を評価することができるため、コミュニケーション能力を測定することもできる。ブログを使った言語能力測定として、荒牧らはブログ・テキストから高齢者の言語能力を評価した [8]。高齢者が理解している語彙数は、若者の約 1.3 倍だという報告がある [9]。その一方で、人間が構文を操る能力は、70 代を後半を境に低下し始めるという報告もある [10]。しかし、実験の結果から、高齢者は、使用する言葉の種類が減る可能性がある。さらに、難易度の高い言葉から使用頻度が減ることが示唆された。また、荻行らは、ブログの面白さを分析した [11]。一般に文章の上手な人間は接続詞を多用せずに読みやすい文章を書くことがわかった。我々が提案するシステムでは、従来の音声での測定に加えて文章や Twitter での測定も可能とした。ブログに書き込むように本システムに文章を入力することで、言語能力を測定することができる。また、既に Twitter などに入力したデータを用いることで言語能力測定への抵抗感の軽減を図る。

## 3. 言語能力測定システム「言秤 WEB」

## 3.1 システムの概要

これまでに我々が開発した言秤 [4] は、直接システムに発話し、音声認識を用いて言語能力測定を行うものである。2017 年の 2 月に京都で行われた認知症相談会において、高齢者に言秤を利用してもらった結果、言秤の利用における問題の一つとして、「測定に対する敷居の高さ」があることがわかった。従来の言秤は、マイクを用いた直接入力を前提としている。一部の高齢者の中には、マイクに向かって話すことに慣れておらず、どこに向かって話せばいいのかわからない様子が見られた。また、村上らは、「音声対話とは一般的に人同士のコミュニケーションに用いるもののため、音声対話システムを利用する際に抵抗感が生じる」と述べている [12]。

そこで、これらの課題を解決するために、利用者のシステム利用の負担を軽減するシステム「言秤 WEB」を開発した。言秤 WEB は WEB ブラウザ上で動くシステムである。言秤 WEB では、従来の音声での測定に加え、テキストデータ、Twitter のツイートデータを用いた言語能力測定を実現する。本システムにおける音声での測定の際は、音声ファイルを用いる。直接システムに向かって話す必要性を無くすことで、抵抗感の軽減を図る。また、音声ファイルを用いることで、利用者が直接測定するのではなく第三者が音声ファイルを用いて測定することも可能にした。テキストデータでの測定は、キーボード等で文字を入力することによる測定である。ただし、入力することそのものに

<sup>†</sup> 和歌山大学大学院システム工学研究科, Graduate School of Systems Engineering, Wakayama University

<sup>††</sup> 奈良先端科学技術大学院大学研究推進機構, Nara Institute of Science and Technology

<sup>†††</sup> 公立諏訪東京理科大学工学部, Faculty of Engineering, Suwa University of Science

<sup>1</sup> [http://www.kanken.or.jp/bjt/survey\\_reports/data/survey\\_reports\\_report01.pdf](http://www.kanken.or.jp/bjt/survey_reports/data/survey_reports_report01.pdf)

表 1: 本研究で用いる言語能力指標

指標	略記	概要
(1) Type・Token 割合	TTR	語彙の量を示す
(2) 頻度・使用者数比	FPU	語彙の特殊性を示す
(3) 日本語学習語彙レベル	JEL	語彙の難しさを示す
(4) 機能表現難度	FNC	文章の難易度を示す
(5) ポライトネス	PLT	文章の丁寧さを示す
(6) 具体性	NER	文章の具体性を示す

手間を感じる人がいる可能性がある。そこで、既に他のシステムで入力されているデータを利用することでより手間の軽減を図る。本システムでは、マイクロブログの1つである Twitter<sup>1</sup> のツイートデータを利用できるようにする。

### 3.2 言語能力指標

本節では、本システムで用いる言語能力指標について述べる。本稿では、言語能力指標として算出されるスコアを「言語能力スコア」と呼ぶ。言語能力指標の一覧を表 1 に示す。これらの指標は、荒牧が加齢による言語能力の変化を調査するために用いたものである [8]。

#### (1) Type・Token 割合 (Type Token Ratio; TTR) :

Type (異なり語数) と Token (延べ語数) の比率 (Type/Token) を示す。この値が大きいほど、語彙量が多いことを意味する。文章全体で集計した値を TTR スコアとする。

#### (2) 頻度・使用者数比 (Frequency per User Popularity; FPU) :

語の特殊性を示す指標である。語の特殊性は、語の出現頻度を語のユーザ数で割った値と定義する。この値が小さいほど一般的な語であり、高いほど特殊な語であることを示す。全単語で算出値を平均した値を FPU スコアとする。

#### (3) 日本語学習レベル (Japanese Education Lexicon Level; JEL) :

語彙の難易度を示す指標である。難易度は日本語学習辞書の語彙レベルを用いた。語彙レベルは 1~6 の 6 段階に分けられ、数字が大きいほど難しい単語となる [14]。単語ごとに語彙レベルを算出し、全単語の平均を JEL スコアとする。

#### (4) 機能表現難度 (Difficulty of Functional Expression; FNC) :

機能表現の難易度を示す。難易度は「日本語機能表現辞書つつじ」[15][16] で設定されている難易度を用いた。この値が大きいほど、文章内で用いられている機能表現の難易度が高いことを示す。文章ごとに算出し、平均した値を FNC スコアとする。

#### (5) ポライトネス (Politeness of Functional Expression; PLT) :

機能表現の丁寧さの度合いを示す。丁寧度は「日本語機能表現辞書つつじ」[15][16] で設定されている丁寧度を用いた。この値が大きいほど、文章内で用い

られている機能表現の丁寧度が高いことを示す。文章ごとに算出し、平均した値を PLT スコアとする。

#### (6) 具体性 (Named Entity Ratio; NER) :

固有名詞の割合を示す。固有名詞の判定は、形態素解析器 JUMAN<sup>2</sup> を用いて行う。文章内の固有名詞、地名、数詞の数を算出し、文章内の全名詞の数で割った値を NER スコアとする。

### 3.3 言語能力測定手法

本節では、言語能力の測定手法として音声データによる測定、文章入力による測定および Twitter による測定について述べる。

#### 3.3.1 音声データによる測定手法

音声データを入力とし、音声認識システム Julius[13] によりテキストへ変換し、言語能力を測定する手法である。Julius はデフォルトで「wav ファイル」および「RAW ファイル」に対応している。しかし、Android や iPhone 等のスマートフォンで録音した音声データは、「amr ファイル」および「m4a ファイル」のため、Julius で音声認識することができない。そこで、音声データを変換可能なフリーソフトである「sox」<sup>3</sup> および「FFmpeg」<sup>4</sup> を用いて、Julius のフォーマットに適した「wav ファイル」へと変換する。変換した音声ファイルを用いて音声認識を行い、テキストを抽出する。抽出したテキストデータをもとに言語能力スコアを算出する。

#### 3.3.2 文章入力による測定手法

テキストボックスに直接文章を入力することで言語能力を測定する手法である。音声データによる測定手法と異なり、音声認識を行う必要がないため、入力されたテキストデータをそのまま用いて、言語能力スコアを算出する。

#### 3.3.3 Twitter データによる測定手法

利用者が Twitter に投稿したツイートから言語能力を測定する手法である。利用者の Twitter アカウントを入力することで利用者のツイートデータ過去 200 件を抽出する。抽出データ数は、Twitter API における 1 回のリクエストでの取得数の上限である 200 件とした。画像のみのツイートやリツイートのようなユーザが発言していないものは除外する。日本人のツイートは、1 ツイート当たり 15 文字程度と

<sup>1</sup><https://twitter.com/>

<sup>2</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

<sup>3</sup><http://sox.sourceforge.net/>

<sup>4</sup><http://www.ffmpeg.org/>

表 2: 発話のテーマ

テーマ
今後、どのような旅行をしてみたいか？
どういう家に住んでみたいか？
アルバイトをしていた時の思い出
旅行の思い出（素晴らしかったこと、失敗談など）
無人島に3つだけ持っていけるとしたら何を持っていくか？
1000万円を1日で使うなら、どんなことに使うか？
学生時代のクラブの思い出

いわれているため<sup>1</sup>、約 3000 文字のデータを用いて言語能力を測定することになる。抽出したツイートデータをもとに言語能力スコアを算出する。

### 3.4 システムの構成

言秤 WEB のシステム構成図を図 1 に示す。本システムは、3.3 節で述べた 3 つの測定方法により言語能力を測定することができる。言秤 WEB の動作の流れを以下に示す。

- (1) 利用者はスマートフォンあるいはパソコン上で言秤 WEB のページにアクセスし、いずれかのデータ（音声データ、テキストデータ、Twitter アカウント）を入力
- (2) 入力されたデータから言語能力スコアを算出
  - (a) 音声データの場合、音声認識モジュールによって音声データからテキストを抽出し、言語処理モジュールによってテキストから各スコアを算出
  - (b) テキストデータの場合、言語処理モジュールによってテキストから各スコアを算出
  - (c) Twitter アカウントの場合、利用者の Twitter アカウントをもとにツイートデータを取得し、言語処理モジュールによってテキスト（ツイート）から各スコアを算出
- (3) 言語能力スコアを送信し、スマートフォンあるいはパソコン上に測定結果を表示

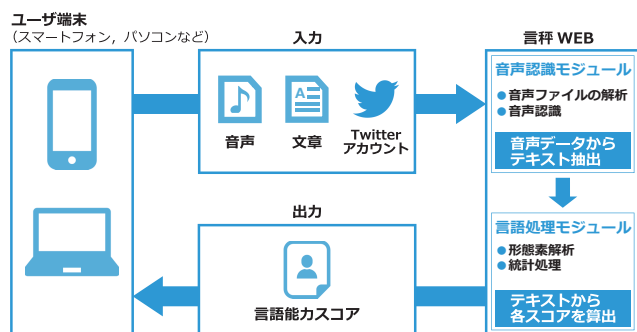


図 1: システム構成図

## 4. 言秤 WEB の利用実験

本実験では、言秤 WEB と言秤を使用してもらい、言語能力の測定結果を表示するまでの操作感を評価する。実験では以下の方法による言語能力の測定を行った。

- 言秤 WEB（音声データ）による測定
- 言秤 WEB（文章入力）による測定
- 言秤 WEB（Twitter データ）による測定
- 従来の言秤による測定

実験協力者によって 4 つの測定方法の実施順序が異なるように設定した。言秤はパソコン（Surface pro4）、言秤 WEB はパソコン（Surface pro4）と Android（ARROWS F-02H）（以下、スマートフォンと表記）の 2 つのデバイスを用いて利用してもらった。このデバイスの実施順番も異なるように設計した。

実験協力者は日本語母語話者である大学生 10 名（男性 5 名、女性 5 名、平均年齢 22.3 才）である。言秤 WEB を用いた音声データおよび文章入力での測定の際には、スピーチのネタ道場を参考に、表 2 に示す 7 つのテーマから選んでもらうことにした。音声データでの測定の場合、スマートフォンに向かって選んだテーマについて 1 分程度発話して音声ファイルを作成してもらい、文章入力での測定の場合、100 文字程度の文章を作成・入力してもらった。

## 5. 結果と考察

表 3 から表 6 に各測定方法の操作感についてのアンケート結果を示す。1 つの測定が終わるごとにアンケートに回答してもらった。言秤 WEB を用いた測定では、パソコン操作についてのアンケートの記入後、スマートフォン操作についてアンケートに記入してもらった。表 3～表 6 における有意確率は、マン・ホイットニーの U 検定によるものである。

### 5.1 測定方法の手間に関して

#### 5.1.1 デバイス別に関する評価

各測定方法の手間に関して質問した結果（表 3（1）～表 5（1））、音声データでの測定に関する評価値の中央値は、パソコンを用いた場合 4、スマートフォンを用いた場合 2 となっており、スコアに有意な差がみられており（ $p=0.025<0.05$ ）、スマートフォン操作の方が手間を感じさせないことが分かった。今回の実験では、パソコンを用いた音声データでの測

<sup>1</sup> <https://blog.twitter.com/official/ja-jp/topics/product/2017/Cramming-Experiment.html>

表 3: 音声データでの測定の操作に関するアンケート

質問項目	デバイス	評価の分布					中央値	最頻値	有意確率
		1	2	3	4	5			
(1) 音声データの測定方法に手間を感じた	パソコン	2	1	1	3	3	4	4,5	0.025*
	スマートフォン	4	3	0	3	0	2	1	
(2) 音声データの測定方法に抵抗感があった	パソコン	4	3	0	3	0	2	1	0.247
	スマートフォン	2	2	1	5	0	3,5	4	
(3) 音声データの測定方法は操作がしやすかった	パソコン	1	4	1	2	2	2,5	2	0.196
	スマートフォン	0	2	1	4	3	4	4	

\*: p&lt;0.05

評価項目：1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する

表 4: 文章入力での測定の操作に関するアンケート

質問項目	デバイス	評価の分布					中央値	最頻値	有意確率
		1	2	3	4	5			
(1) 文章入力の測定方法に手間を感じた	パソコン	7	3	0	0	0	1	1	0.101
	スマートフォン	4	3	1	0	2	2	1	
(2) 文章入力の測定方法に抵抗感があった	パソコン	8	2	0	0	0	1	1	0.018*
	スマートフォン	6	1	0	2	1	1	1	
(3) 文章入力の測定方法は操作がしやすかった	パソコン	0	0	0	2	8	5	5	0.224
	スマートフォン	0	1	2	4	3	4	4	

\*: p&lt;0.05

評価項目：1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する

表 5: Twitter での測定の操作に関するアンケート

質問項目	デバイス	評価の分布					中央値	最頻値	有意確率
		1	2	3	4	5			
(1) Twitter の測定方法に手間を感じた	パソコン	9	1	0	0	0	1	1	0.091
	スマートフォン	5	4	0	1	0	1.5	1	
(2) Twitter の測定方法に抵抗感があった	パソコン	7	2	0	1	0	1	1	1.000
	スマートフォン	7	2	0	1	0	1	1	
(3) Twitter の測定方法は操作がしやすかった	パソコン	0	0	0	2	8	5	5	0.260
	スマートフォン	0	2	0	2	6	5	5	

\*: p&lt;0.05 評価項目：1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する

定に関しても、スマートフォンで録音した音声データを用いて測定を行ってもらった。実際の利用用途として、「簡単に音声録音できるデバイス（スマートフォンなど）を用いて音声録音し、そのデータを用いて言語能力を測定する」という状況を想定しているため、できる限り利用状況に合わせた実験を行った。自由記述において、「スマートフォンで録音した音声データをパソコンに送信する作業に抵抗感を感じた」という意見が得られ、音声データでの測定に関してはスマートフォン操作の方が評価が高かった。文章入力での測定に関する評価値の中央値は、パソコンを用いた場合 1、スマートフォンを用いた場合 2 となっていた。また、Twitter での測定に関する評価値の中央値は、パソコン・スマートフォンのいずれも 1 となっていた。文章入力および Twitter での測定は共に評価が高かったが、有意な差はみられなかった（文章入力：p=0.101, Twitter：p=0.091）。普段から日常的にパソコンおよびスマートフォンに慣れている人が多かったため、自由記述において、「日常的に触っているから文字を入力することに慣れている」という意見が得られた。デバイスに関係なく、測定方法に手間を与えな

いことが分かった。

### 5.1.2 言評と言評 WEB に関する評価

言評による測定の手間に関して質問した結果（表 6（1））、評価値の中央値は、音声データの場合 4、文章入力の場合 1、Twitter の場合 1 および言評の場合 2 となっており、Twitter での測定と言評の測定との間にのみ有意な差がみられた（音声データ：p=0.549, 文章入力：p=0.079, Twitter：p=0.016<0.05）。言評に関する自由記述において、「そもそも声を出すのが面倒」という意見が得られた。今回のシステムで実装した文章入力での測定と Twitter での測定という 2 つの機能を用いることで、言評よりも測定の手間を軽減できる可能性がある。

## 5.2 測定方法の抵抗感に関する評価

### 5.2.1 デバイス別に関する評価

各測定方法の抵抗感に関して質問した結果（表 3（2）～表 5（2））、文章入力での測定に関する評価値の中央値は、パソコン・スマートフォンのいずれも評価値の中央値は 1

表 6: 言秤での測定の操作に関するアンケート

質問項目	評価の分布					中央値	最頻値	言秤 WEB との有義確率		
	1	2	3	4	5			音声データ	文章入力	Twitter
(1) 言秤の測定方法に手間を感じた	4	2	0	3	1	2	1	0.549	0.079	<b>0.016*</b>
(2) 言秤の測定方法に抵抗感があった	3	1	1	4	1	3.5	4	0.321	<b>0.009*</b>	<b>0.037*</b>
(3) 言秤の測定方法は操作がしやすかった	0	0	1	2	7	5	5	<b>0.009*</b>	0.549	0.549

\* :  $p < 0.05$

有意確率は、言秤 WEB をパソコンで測定した際の評価との有意確率となっている。

評価項目：1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する

であったが、パソコンは3以上の評価をした実験協力者がいなかったのに対し、スマートフォンでは3以上の評価をした実験協力者が3名おり、評価結果に有意な差がみられた ( $p=0.018 < 0.05$ )。自由記述では、「スマートフォンで長い文章を書くのは疲れる」という意見があり、長い文章の入力においてはパソコン操作の方が抵抗感が少ない可能性がある。音声データでの測定に関する評価値の中央値は、パソコンを用いた場合2、スマートフォンを用いた場合3.5となっており、有意な差はみられなかった ( $p=0.247$ )。また、Twitterでの測定に関しては、評価が完全に一致していたため、有意な差はみられなかった ( $p=1.00$ )。これらの結果より、音声データでの測定とTwitterでの測定において、デバイスによる大きな差はないと考えられる。

### 5.2.2 言秤と言秤 WEB に関する評価

言秤による測定の抵抗感に関して質問した結果(表6(2)), 評価値の中央値は、3.5となり、文章入力(評価値:1)およびTwitterでの測定(評価値:1)との間に有意な差がみられた(音声データ:  $p=0.321$ , 文章:  $p=0.009 < 0.05$ , Twitter:  $p=0.037 < 0.05$ )。言秤に関する自由記述において、「パソコンに向かって話すのが苦痛」および「パソコンに向かって話すことが恥ずかしい」という意見が得られた。そのため、文章での測定とTwitterでの測定は、言秤よりも抵抗感を軽減できる可能性がある。

### 5.3 測定方法の操作に関する評価

#### 5.3.1 デバイス別に関する評価

各測定方法の操作に関して質問した結果(表3(3)~表5(3)), 音声データでの測定に関する評価値の中央値は、パソコンを用いた場合2.5、スマートフォンを用いた場合4となっており、文章入力では、パソコンを用いた場合5、スマートフォンを用いた場合4となっており、Twitterでは、パソコンを用いた場合5、スマートフォンを用いた場合5となっていた。評価の結果には差がみられたが、全ての測定において、有意な差はみられなかった(音声データ:  $p=0.196$ , 文章:  $p=0.224$ , Twitter:  $p=0.260$ )。音声データでの測定に関する自由記述において、「スマートフォンの方が測定してそのまま送信するだけだから楽」という意見が得られたが、有意な差は見られなかった。文章入力およびTwitterの自由記述において、「文字を打つだけなので楽」という意見が得られたため、デバイスにおける差異はみられないことがわかった。

### 5.3.2 言秤と言秤 WEB に関する評価

言秤による測定の操作に関して質問した結果(表6(3)), 評価値の中央値は5となり、音声での測定(評価値:2.5)との間にのみ有意な差がみられた(音声データ:  $p=0.009 < 0.05$ , 文章入力:  $p=0.549$ , Twitter:  $p=0.549$ )。言秤に関する自由記述において、「ボタンをクリックして話すだけなので操作はしやすかった」という意見が得られ、また、言秤 WEB の音声データでの測定よりも高評価が得られた。これは、言秤が直接パソコンに入力できることが原因だと考えられる。言秤 WEB はパソコンでの直接入力に対応しておらず、言秤 WEB の自由記述においても、「パソコンで直接入力できるようにしてほしい」という意見が得られた。そのため、今後 WEB 上でも直接入力できるように改善していく。

### 5.4 4種類の測定方法に関して

従来システムでの問題点を解決するために、提案システムでは音声データでの測定に加えて、文章およびTwitterでの測定の2つの測定方法を利用可能にした。音声データ、文章入力、Twitter、言秤の4つの測定に関して、「手間を感じた」、「抵抗感があった」、「操作がしやすかった」という点について、順位付けをしてもらった。順位付け結果の度数分布を表7から表9に示す。表7、表8に関しては、合計点数の最大が40点であり、総和が高いほど総合的評価が良いと判断でき、表9に関しては、合計点数の最小が10点であり、総和が低いほど総合的評価が良いと判断できる。

表7から表9より、Twitterでの測定が最も評価が高かった。総和の総合評価は、手間を感じた:37点、抵抗感があった:38点および操作がしやすかった:10点と1番高かった。これは、Twitterでの測定では、自身のアカウントを入力するだけで測定できるためであると考えられ、自由記述においても、「ユーザ名を入力するだけだから一番簡単」という意見が多く見られた。一方で、「Twitterのアカウントを利用することによる、個人情報の取得に不安を感じる」「何気ないことを書き込むことが多いため、文章をきちんと書く(文章入力での測定)よりは話しの丁寧さなどが下がる」という意見も見られた。今後、Twitterに投稿されるツイートから正確な言語能力測定が可能なのかを分析する必要がある。

また、言秤と比較して、文章入力での測定も評価が高かった。総和の総合評価は、手間を感じた:27点、抵抗感があった:29点および操作がしやすかった:29点と2番目に高かった。そのため、従来よりも抵抗感を軽減できたと考え

表 7: 4 種類の測定方法に関する手間を感じた順

	音声データの測定	文章入力での測定	Twitter の測定	言評
ランク 1	3	2	1	4
ランク 2	4	0	0	6
ランク 3	3	7	0	0
ランク 4	0	1	9	0
総和	20	27	37	16

- ・総和は、出現頻度と順位とを乗算した結果の総和である。
- ・ランクは、4 段階評価の順位である。
- ・総和が高いほど総合的評価が良い。

表 8: 4 種類の測定方法に関する抵抗感があつた順

	音声データの測定	文章入力での測定	Twitter の測定	言評
ランク 1	5	1	0	4
ランク 2	4	1	0	5
ランク 3	1	6	2	1
ランク 4	0	2	8	0
総和	16	29	38	17

- ・総和は、出現頻度と順位とを乗算した結果の総和である。
- ・ランクは、4 段階評価の順位である。
- ・総和が高いほど総合的評価が良い。

表 9: 4 種類の測定方法に関する操作がしやすかった順

	音声データの測定	文章入力での測定	Twitter の測定	言評
ランク 1	0	0	10	0
ランク 2	2	5	0	3
ランク 3	3	1	0	6
ランク 4	5	4	0	1
総和	33	29	10	28

- ・総和は、出現頻度と順位とを乗算した結果の総和である。
- ・ランクは、4 段階評価の順位である。
- ・総和が低いほど総合的評価が良い。

表 10: 操作がしやすかったデバイスについてのアンケート

測定手法	パソコン (人)	スマートフォン (人)
音声データでの測定	4	6
文章入力での測定	10	0
Twitter で測定	7	3

られる。文章入力での測定は、テキストボックスに文字を書くだけであり、抵抗感を感じる人は少なかった。なお、自由記述においては、「Twitter での測定と比較して、文字をいっぱい入力するので大変だった」という意見も見られた。

音声データでの測定に関しては、評価が低かった。総和の総合評価は、手間を感じた：20 点、抵抗感があつた：16 点および操作がしやすかった：33 点となった。また、言評での測定に関しても、音声データでの測定と同様に評価が低く、総和の総合評価は、手間を感じた：16 点、抵抗感があつた：17 点および操作がしやすかった：28 点となった。音声データでの測定と同じように、「パソコンに向かって話すことをほとんどしたことがないので、抵抗感を感じる」という意見が多く見られた。また、言評は言語能力測定において 3 回音声の録音を行う。その 3 回音声を録音する作業に手間を感じる人も多かった。しかし、操作方法に関しては、「クリックするだけなので操作は簡単」という意見が多く見られた。

したがって、Twitter での測定を用いることにより言語能

力測定に間して、手間や抵抗感を軽減できる可能性がある。Twitter での測定に不安を感じる場合は、文章入力での測定を用いることで、Twitter での測定と同様に手間や抵抗感を軽減できると考えられる。今回の実験では、実験中に音声データや文章を作成してもらった。事前に音声データや文章を用意してもらい、そのデータを用いて言語能力測定を行うことで、音声データでの測定も手間や抵抗感を軽減できる可能性がある。

## 5.5 各デバイスでの測定に関して

表 10 に各デバイスでの操作感に関するアンケート結果を示す。表中の数値は回答した人数である。各測定手法に関して、操作がしやすかったデバイスについて一対比較を行った結果、音声データでの測定に関しては、パソコン操作選んだ人が 4 人、スマートフォン操作を選んだ人が 6 人とわずかに多い結果が得られた。一方で、文章入力での測定に関しては、全員がパソコン操作の方が操作を選び、Twitter での測定に関しては、パソコン操作選んだ人が 7 人、スマートフォン操作を選んだ人が 3 人とパソコン操作を選ぶ人が多かった。そのため、スマートフォンの方が音声データでの測定の場合手間を軽減でき、パソコン操作の方が文章入力での測定の場合操作がしやすい可能性がある。

理由としては以下の 2 点が考えられる。

- 実験協力者が全員 iPhone ユーザである
- スマートフォンは長い文章の入力に不向きである

今回、実験するにあたって日常利用しているスマートフォンの機種に関するアンケートに回答してもらった。その結果、全てのユーザが iPhone ユーザだった。今回の実験で用いた Android の操作に慣れていないため、自由記述において「Android 操作がわからなかった」「テンキー入りに慣れていない」という意見が多く見られた。また、文章入力での測定において、スマートフォンで長文を記入することに手間、抵抗感を感じると答えた人が多かった。

## 6. おわりに

本研究では、これまでに開発してきた言評の問題点を解決するために、測定負担の軽減を目的とした「言評 WEB」を開発した。開発システムは、音声データ、文章入力および Twitter のツイートデータを用いて言語能力を測定することで、測定の負担の軽減を図っている。開発システムを用いた実験により、以下の 2 点を明らかにした。

- (1) Twitter と文章入力での測定は言語能力測定の抵抗感を軽減できる可能性がある。
- (2) 音声データでの測定に関してはスマートフォン、それ以外の測定に関してはパソコンの方が操作しやすい。

本システムでは、音声による測定の抵抗感を改善することは出来なかった。ただし、文章入力、Twitter データでの測定により、抵抗感を軽減できる可能性がある。また、デバイスごとに操作しやすい測定方法があることがわかった。ただし、今回は大学生を対象としての評価を行ったため、今後は高齢者などによるシステムの操作性についての評価を行う。

## 参考文献

- [1] 北村育子：特別養護老人ホームに暮らす認知症利用者のアクセスメントの実際とその重症度評価における主要評価項目，日本福祉大学社会福祉論集，pp.33-46（2006）。
- [2] 平成 29 年度外国人留学生在籍状況調査などについて，独立行政法人日本学生支援機構，[http://www.jasso.go.jp/about/statistics/intl\\_student/data\\_2017.html](http://www.jasso.go.jp/about/statistics/intl_student/data_2017.html)（参照 2018 年 2 月 6 日）
- [3] 山口壘：日本企業における留学生人材の活用と労働市場での位置づけ，法政大学比較経済研究所，pp.1-21（2016）。
- [4] 宮部真衣，四方朱子，久保圭，荒牧英治：音声認識による認知症・発達障害スクリーニングは可能か？— 言語能力測定システム“言秤”の提案—，ワークショップ 2014 (GN Workshop 2014) 論文集，Vol. 2014，pp.1-8（2014）。
- [5] Emiddia, L., Pietro, S., Alessandra, F., Leslie, R. : Language and social competence in typically developing children and late talkers between 18 and 35 months of age , Early Child Development and Care, 186(3) , pp.436-452（2015）。
- [6] Christine, Y., Allison L. , Diane K. , Albert L. : Language of Early- and Later-identified Children With Hearing Loss, Pediatrics, Vol.102, No.5 , pp.1161-1171（1998）。
- [7] 鳥飼玖美子：真のコミュニケーション能力を培う為に— 母語と外国語を繋ぐ言語教育 —，日本学術会議「学術の動向」，13 卷 1 号，pp.56-58（2006）。
- [8] 荒牧英治，久保圭，四方朱子：老いとくことば>：ブログ・テキストから測る老化，情報処理学会研究報告，Vol. 2014-DBS-159, No.23, pp.1-6（2014）。
- [9] 呉田陽一，伏見貴夫，佐久間尚子：言語能力の加齢変化，第 9 回東京老年学会誌，9 卷，pp.200-205（2003）。
- [10] Kemper, S., Marquis, J. and Thompson, M. : Longitudinal change in language production : effects of aging and dementia on grammatical complexity and propositional content, Psychology and Aging, Vol.16, No.4, pp.600-614（2001）。
- [11] 萩行正嗣，柴田知秀，黒橋禎夫：表層・語彙的特徴量に基づくブログの面白さ分析，情報処理学会研究報告，Vol.2009-NL-185, pp.45-52（2008）。
- [12] 村上拓也，李晃伸，西川由理，小島良宏，遠藤充：音声対話システムにおけるシステムからの話しかけと他者性認知の関連性の調査，HAI シンポジウム 2015，pp.238-243（2015）。
- [13] 河原達也，李晃伸：音声認識ソフトウェア Julius，人工知能学会誌，Vol.20, No.1, pp.41-49（2005）。
- [14] 砂川有里子：学習辞書編集支援データベース作成について-『学習辞書科研』プロジェクトの紹介-，日本語教育連絡会議論文集，Vol.24, pp.164-169（2012）。
- [15] 松吉俊，佐藤理史，宇津呂武仁：日本語機能表現辞書の編纂，自然言語処理，Vol.14, No.5, pp.123-146（2007）。
- [16] 松吉俊，佐藤理史：文体と難易度を制御可能な日本語機能表現の言い換え，自然言語処理，Vol.15, No.2, pp.75-99（2008）。